

FIG. 1-1

ESX ORF DNA Sequence (1 to 1116) → 1-phase Translation
 DNA sequence 1116 b.p. ATGGCTGCAACC ... agtccgaactgaa linear

371 Amino acids
 MW: 41428 Dalton

<p>1/1 ATG GCT GCA ACC TGT GAG ATT AGC AAC ATT TTT AGC AAC TAC TTC AGT GCG ATG TAC AGC TCG GAG GAC TCC ACC CTG GCC TCT GTT CCC Met ala ala thr cys glu ile ser asn ile phe ser asn tyr phe ser ala met tyr ser ser glu asp ser ser glu asp ser thr leu ala ser val pro 91/31</p> <p>CCT GCT GCC ACC TTT GGG GCC GAT GAC TTG GTA CTG ACC CTG AGC AAC CCC CAG ATG TCA TTG GAG GGT ACA GAG AAG GCC AGC TGG TTG pro ala ala thr phe gly ala asp asp leu val leu thr leu ser asn pro gln met ser leu glu gly thr glu lys ala ser trp leu 181/61</p> <p>GGG GAA CAG CCC CAG TTC TGG TCG AAG ACG CAG GTT CTG GAC TGG ATC AGC TAC CAA GTG GAG AAG AAC AAG TAC GAC GCA AGC GCC ATT gly glu gln pro gln phe trp ser lys thr gln val val asp trp ile ser tyr gln val glu lys asn lys tyr asp ala ser ala ile 271/91</p> <p>GAC TTC TCA CGA TGT GAC ATG GAT GGC ACC CTC TGC AAT TGT GCC CTT GAG CTG CGT CTG GTC TTT GGG CCT CTG GGG GAC CAA asp phe ser arg cys asp met asp gly ala thr leu cys asn cys ala leu glu glu leu arg leu val phe gly pro leu gly asp gln 361/121</p> <p>CTC CAT GCC CAG CTC CGA GAC CTC ACT TCC AGC TCT TCT GAT GAG CTC AGT TGG ATC ATT GAG CTG CTG GAG AAG GAT GGC ATG GCC TTC Iau his ala gln leu arg asp leu thr ser ser asp glu leu ser ser trp ile ile gln glu leu leu asp asp gln gln ala ser pro tyr his 451/151</p> <p>CAG GAG GCC CTA GAC CCA GGG CCC TTT GAC CAG GGC AGC CCC TTT GCC CAG GAG CTG CTG GAC GAC GGT CAG CAA GCC AGC CCC TAC CAC gln glu ala leu asp pro gly pro phe asp gln gly ser pro phe ala gln glu leu leu asp asp gln gln ala ser pro tyr his 541/181</p> <p>CCC GGC AGC TGT GGC GCA GGA GCC CCC TCC CCT GGC AGC TCT GAC GTC TCC ACC GCA GGG ACT GGT GCT TCT CGG AGC TCC CAC TCC TCA pro gly ser cys gly ala gly ala pro ser pro gly ser ser asp val ser asp val ser thr ala gly thr gly ala ser arg ser ser his ser ser 631/211</p> <p>GAC TCC GGT GGA AGT GAC GTG GAC CTG GAT GGC AAG CTC TTC CCC AGC GAT GGT TTT CGT GAC TGC AAG AAG GGG GAT CCC asp ser gly gly ser asp val asp leu asp pro thr asp gly lys leu phe pro ser asp gly phe arg asp cys lys lys gly asp pro</p>	<p>31/11 121/41 211/71 301/101 331/111 421/141 511/171 601/201 691/231</p> <p>61/21 151/51 241/81 331/111 421/141 511/171 601/201 691/231</p>
---	--



FIG. 1-2

811/271 AGA GGC ACC CAC CTG TGG GAG TTC ATC CGG GAC ATC CTC ATC CAC CCG GAG CTC AAC GAG CTC ATG AAG TGG GAG ATC CGG CAT GAA
arg gly thr his leu trp glu phe ile arg asp ile leu ile his pro glu leu asn glu gly leu met lys trp glu asn arg his glu
901/301 931/311 961/321
841/281 GGC GTC TTC aag TTC CTG CGC TCC GAG GCT GTG GCC CAA CTA TGG GGC CAA AAG AAA AAG AAC ATG ACC TAC GAG AAG CTG AGC
gly val phe lys phe leu arg ser glu ala val ala val ala gln leu trp gly gln lys lys asn ser asn met thr tyr glu lys leu ser
991/331 1021/341 1051/351
CGG GGC ATG AGG TAC TAC TAC AAA CGG GAG ATC CTG GAA CGG GTG GAT GGC CGG CGA CTC GTC TAC AAG TTT GGC AAA AAC TCA AGC GGC
arg ala met arg tyr tyr tyr lys arg glu ile leu glu arg val asp gly arg arg leu val tyr lys phe gly lys asn ser ser gly
1081/361 1111/371
TGG AAG GAG GAA GAG GTT CTC CAG AGT CGG AAC TGA
trp lys glu glu glu val leu gln ser arg asn OPA





FIG. 2a

MAATCEISNIFS NYFSAMYSS EDSTLASVPPAATFGADDVLTL SNPQMSLEG	53
TEKASWLGEQPQFW SKTQVLDW ISYQVEKNKYDASAIDFSRCMDGATLCNCA	106
LEELRLVFGPLGDQLHAQLRDLTSSSDELSWIIELLEKDGM AFQEALDPGP F	159
DQGSPFAQEELLDDGQQASPYHPGSCGAGAPSPGSSDVSTAGTGASRSSHSSDS	212
GGSDVDLDPTDGKLFP SDGFRDCKKGDPKHGKRKRGRPRKLSKEYWDCLEGKK	265
SKHAPRGTHLWEFIRDILIHPELMKWENRHEGVFKFLRSEAVAQLWGQK	318
KKNSNMTYEKLSRAM RYY KREILERVDGRRLVYKFGKNSSGWKEEEVL QSRN	371



FIG. 2b

Consensus⁺:
ESX (64-103):
P W V W W E
PQFWSKTQVLDWISYQVEKNNKYDASAIDFSRCDMGDATLC
P+ W++T V DW+ + V N++ +DF + M+GA LC
PRQWTETHVRDWWMMWAV--NEFSLKGVDQKFCMNGAAC

FIG. 2c

ESX (188-238):

APS PGSSDVSTAGTGASRSSHSSDSGGSDV DPTDGKLFPSDGFRDCCKG
APS S S++ + S SS S S SD + + L PS F G
APSSAPSHASSASSSSSSSSSSDEFDDLLNPPSSNFESMSLG

FIG. 2d

Consensus⁺:
LWQFLLL D
* •* • * *

α_1 β_1 β_2 α_2 "turn" α_3 β_3 "wing" β_4

HLWEFIRDILIHPELMKWNENRHEGVFKFLRSEAVAQLWGQKKKKNSNMITYEKLSRAMRRYYKREIILERVGRRRLVYKF
+LWEF+ +L +KW R +G+FK + S+AV++LWG+ K +M YE + RA+RYYY+R IL +V+G+RLVY+F
YLWEFLALLQDKATCPKYIKWTQREKGIFKLVLDSKAWSRLWGKHKNKPDMNYETMGRALRYYQRGILAKVEGQRLVYQF

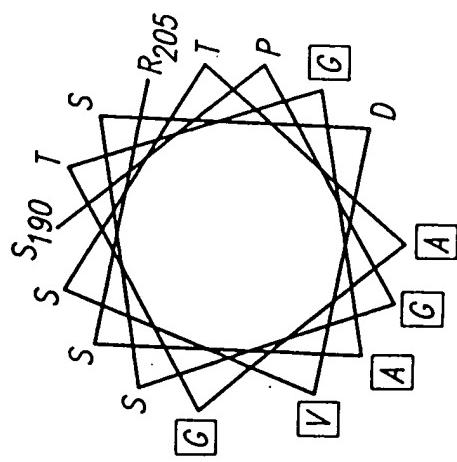




FIG. 2e

MAATCEISNIFS NYFSAMYSS EDSTLASVPPAATFGADDVLTLSNPQMSLEG	53
TEKASWLGEQPQFWSKTQVLDWISYQVEKNKYDASAIDFSRCMDGATLCNCA	106
LEELRLVFGPLGDQLHAQLRDLTSSSSDELSWI ELLKDGM AFQEALDPGPF	159
DQGSPFAQELLDDGQQASPYHPGSCGAGAPSPGSSDVSTAGTGASRSSHSSDS	212
GGSDVDLDPTDGKLFP SDGFRDCKKGDPKHGKRKRPRKLSKEYWDCLEGKK	265
SKHAPRGTHLWEFIRDILIHPELNEGLMKWENRHEGVFKFLRSEAVAQLWGQK	318
KKNSNMTYEKLSRAMRYYKREILERVDGRRLVYKFGKNSSGWKEEEVLQSRN	371

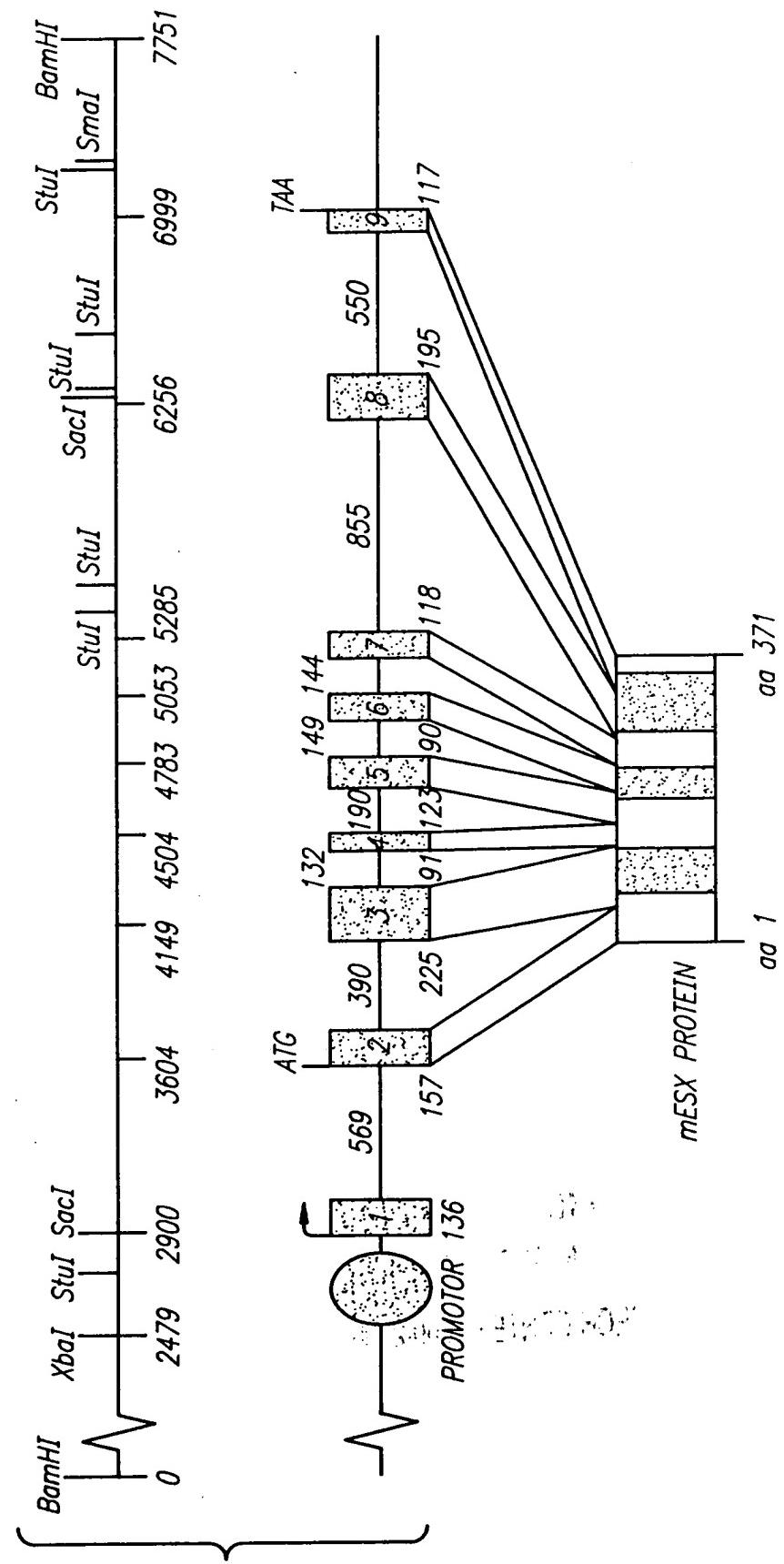


FIG. 3

FIG. 4-1

1/1 ATG GCT GCA ACC TGT GAG ATT AGC AAC ATT TTT AGC AAC TAC TTC AGT GCG ATG TAC AGC
Met ala ala thr cys glu ile ser asn ile phe ser asn tyr phe ser ala met tyr ser
61/21 TCG GAG GAC TCC ACC CTG GCC TCT GTT CCC CCT GCT GCC ACC TTT GGG GCC GAT GAC TTG
ser glu asp ser thr leu ala ser val pro pro ala ala thr phe gly ala asp asp leu
121/41 GTA CTG ACC CTG AGC AAC CCC CAG ATG TCA TTG GAG GGT ACA GAG AAG GCC AGC TGG TTG
val leu thr leu ser asn pro gln met ser leu glu gly thr glu lys ala ser trp leu
181/61 GGG GAA CAG CCC CAG TTC TGG TCG AAG ACG CAG GTT CTG GAC TGG ATC AGC TAC CAA GTG
gly glu gln pro gln phe trp ser lys thr gln val leu asp trp ile ser tyr gln val
241/81 GAG AAC AAG TAC GAC GCA AGC GCC ATT GAC TTC TCA CGA TGT GAC ATG GAT GGC GCC
glu lys asn lys tyr asp ala ser ala ile asp phe ser arg cys asp met asp gly ala
301/101 331/111 ACC CTC TGC AAT TGT GCC CTT GAG GAG CTG CGT CTG GTC TTT GGG CCT CTG GGG GAC CAA
thr leu cys asn cys ala leu glu glu leu arg leu val phe gly pro leu gly asp gln
361/121 E3 E4 391/131 CTC CAT GCC CAG CTG CGA GAC CTC ACT TCC AGC TCT TCT GAT GAG CTC AGT TGG ATC ATT
leu his ala gln leu arg asp leu thr ser ser ser ser asp glu leu ser trp ile ile
421/141 451/151 E4 E5 GAG CTG GAG AAG GAT GGC ATG GCC TTC CAG GAG GCC CTA GAC CCA GGG CCC TTT GAC
glu leu glu lys asp gly met ala phe gln glu ala leu asp pro gly pro phe asp



FIG. 4-2

481/161 CAG GGC AGC CCC TTT GCC CAG GAG CTG CTG GAC GAC GGT CAG CAA GCC AGC CCC TAC CAC
gln gly ser pro phe ala gln glu leu leu asp asp gly gln gln ala ser pro tyr his
541/181 CCC GGC AGC TGT GGC GCA GGA GCC Ccc TCC CCT GGC AGC TCT GAC GTC TCC ACC GCA GGG
pro gly ser cys gly ala gly ala pro ser pro gly ser ser asp val ser thr ala gly
601/201 ACT GGT GCT TCT CGG AGC TCC CAC TCC TCA GAC TCC GGT GGA AGT GAC GTG GAC CTG GAT
thr gly ala ser arg ser ser his ser asp ser gly gly ser asp val asp val asp leu asp
661/221 CCC ACT GAT GGC AAG CTC TTC CCC AGC GAT GGT TTT CGT GAC TGC AAG AAG GGG GAT CCC
pro thr asp gly lys leu phe pro ser asp gly phe arg asp cys lys lys gly asp pro
721/241 AAG CAC GGG AAG CGG AAA CGA CGG CGG CCC CGA AAG CTG AGC AAA GAG TAC TGG GAC TGT
lys his gly lys arg gly arg pro arg lys leu ser lys glu tyr trp asp cys
781/261 CTC GAG GGC AAG AAG AAC CAC GCG CCC AGA GGC ACC CAC CTG TGG GAG TTC ATC CGG
leu glu gly lys lys ser lys his ala pro arg gly thr his leu trp glu phe ile arg
841/281 GAC ATC CTC ATC CAC CCG GAG CTC AAC GAG GGC CTC ATG AAG TGG GAG AAT CGG CAT GAA
asp ile leu ile his pro glu leu asn glu gly leu met lys trp glu asn arg his glu
901/301 GGC GTC TTC aag TTC CTG CGC TCC GAG GCT GTG GCC CAA CTA TGG GGC CAA AAG AAA AAG
gly val phe leu arg ser glu ala val ala gln leu trp gly gln lys lys lys

511/171 E5 ↓ E6
631/211 E6 ↓ E7 691/231
751/251 E6 ↓ E7 811/271
871/291 E7 ↓ E8
931/311

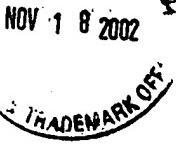




FIG. 4-3

961/321 AAC AGC AAC ATG ACC TAC GAG AAG CTG AGC CGG GCC ATG AGG TAC TAC AAA CGG GAG
asn ser asn met thr tyr glu lys leu ser arg ala met arg tyr tyr lys arg glu
1021/341 ATC CTG GAA CGG GTG GAT GGC CGG CGA CTC GTC TAC AAG TTT GGC AAA AAC TCA AGC GGC
ile leu glu arg val asp gly arg arg leu val tyr lys phe gly lys asn ser ser gly
1081/361 TGG AAG GAG GAA GAG GTT CTC CAG AGT CGG AAC TGA
trp lys glu glu glu val leu gln ser arg asn OPA

991/331 E8 E9



FIG. 5-1

mESX	1 MAATCEISNVFSNYFNAMYSSEDPTLAPAPP.TTFGTEDLVLTLNNQQMT	49
	: . : . .	
hESX	1 MAATCEISNIFSNYFSAMYSSEDSTLASVPPAATFGADDLVLTLSNPQMS	50

mESX 100 TLCSCALEELRLVFGPLGDQLHAQLRDLTSNSSDELSWIIELLEKDGMF 149

 hESX 101 TLCNCALEELRLVFGPLGDQLHAQLRDLTSSSSDELSWIIELLEKDGMF 150

mESX 150 QESLGDLGSPSDQGSPFAQEELLDDGRQASPYCSTYGP_GAPSPGSSDVSTA 199
hESX 151 QEAL.DPGPFDQGSPFAQEELLDDGQQASPYHPGSCGAGAPSPGSSDVSTA 199

The diagram shows the sequence alignment of mESX and hESX proteins. The mESX sequence starts at position 150 and ends at 199. The hESX sequence starts at position 151 and ends at 199. A vertical arrow labeled E_4/E_5 points down to the sequence between positions 175 and 176. Another vertical arrow labeled E_5/E_6 points down to the sequence between positions 194 and 195. Conserved amino acid residues are highlighted with vertical bars above the sequences. The E_4/E_5 site is located in a highly conserved region, while the E_5/E_6 site is located in a less conserved region.

The figure shows two protein sequences aligned horizontally. The top sequence is mESX 200 and the bottom sequence is hESX 200. A vertical arrow labeled E_6/E_7 points to a specific motif in the mESX sequence. The motif consists of a short stretch of hydrophobic amino acids (I, L, D) followed by a highly conserved sequence: **KKGEPKHGKRKGR**. In the hESX sequence, this motif is preceded by a different set of hydrophobic residues (G, S, R, H, S, S). The alignment uses standard sequence comparison symbols: dots for matches, colons for conservative substitutions, and dashes for gaps.

mESX 250 PRKLSKEYWDCLEGKKSKHAPRGTHLWEFIRDILIHPELNEGLMKWENRH

hESX 250 PRKLSKEYWDCLEGKKSKHAPRGTHLWEFIRDILIHPELNEGLMKWENRH



FIG. 5-2

mESX 300 [EGVFKFLRSEAVAQLWGQKKNSNMTYEKLSRAMRYYYKREILERVDGRR] 349
hESX 300 [EGVFKFLRSEAVAQLWGQKKNSNMTYEKLSRAMRYYYKREILERVDGRR] 349

mESX 350 [LVYKFGKNSSGWKEEEVGESRN] 371
hESX 350 [LVYKFGKNSSGWKEEEVLQSRN] 371



POINTED/A-Region



Serine-Rich Box



Nuclear Targeting Sequence



ETS-DNA Binding Domain



FIG. 6

mESX -347	TCAGCCCTGGCCAGGCCCGCCAGGAAGAATTCCAGGGCCAGAGGGCAGCC	-298
hESX -350	CCATCTCTGGCCTGGCCCCCTGGGAGGAATTCCCTGGGCCAGAGGGCAGCC	-301
mESX -297	TAAGGCACAGATGCCACCCCTGCAATGTTCCCGCCACCTGCCAGTTCA	-246
hESX -300	GAAAGCACAGATGCCACCCAGCAACGTTCCCGCCACCTGCCAGGCCA	-251
mESX -247	GTACCCAGGGCCCACCCAGAGGGTGCAGAACAGATTCTGACAATC	-198
hESX -250	GTGCCCGTGCCACCCAGAGGGTGCAGAACAGACTCTGACAATC	-201
mESX -197	ATTAACCAGCCAGGCCTGATTTCCAGCACCGCCTGCTAGGATCTGGC	-146
hESX -200	ATTAACCAGCCGGGCCTGATTTCCAGCACCTGCCTGCTAAGATCCGGC	-151
mESX -147	CAAGTGGCACGGAATATGCAAATCACCTGGGACAGGGAGGCCAGTCTGAA	-98
hESX -150	CAAGTGGCACTGAATATGCAAATCACCTGGGGCCAGGAGGCCAGTCTAAA	-101
mESX -97	GGCCAGGAAATCCCCAGCATGAAAGGCCACCAAGCTCAGGTTACAACCG	-46
hESX -100	GGCCAGGAAATCCCCCTCCATGAAAGACACCAAGCTCAGGTTACTGCAG	-49
mESX -47	GGGACGTACGCCAAGACCTG...GAGGGGAGGGAGCTCTGCTTTG	-1
hESX -50	GGGACACACTATAAGCCCTGAGCTCAGGGAGGAGCTCCCTCCAGG	-1
mESX +1	ATAGAGCGGGTGGGGCAGCGCCCTGGCCACACTCATCACTGCTACCT	+50
hESX +1	ATAGAGCCGGGTAGGGGAGCGCAGCGGCCAGATACCTCAGCGCTACCT	+50

Ets	extended palindrome	USF
AP-2	SPI/GC box	Oct
NF-kB	CCAAT box & Inr element	

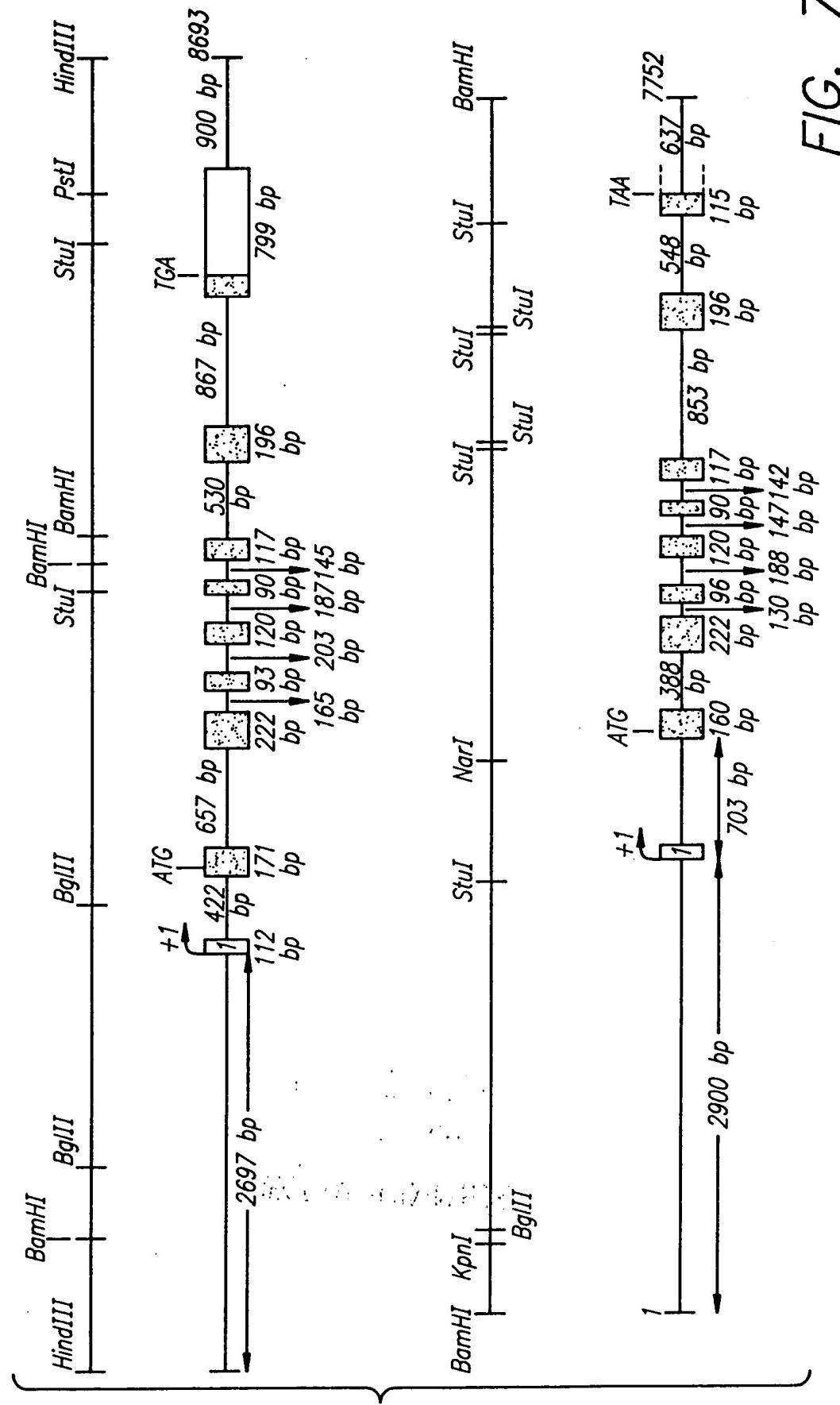


FIG. 7



WT 5' GGAGGAGGGCTGCTTGAGGAAGTATAAGAAT 3'
m1 5' - - - - T A - - - - 3'
m2 5' - - - - - C - - - - 3'
m3 5' - - - - - A G - - - - 3'
m4 5' - - - - - C C - - - - 3'
m5 5' - - - - - C - - - - 3'

Fig. 8A

100-11001-03

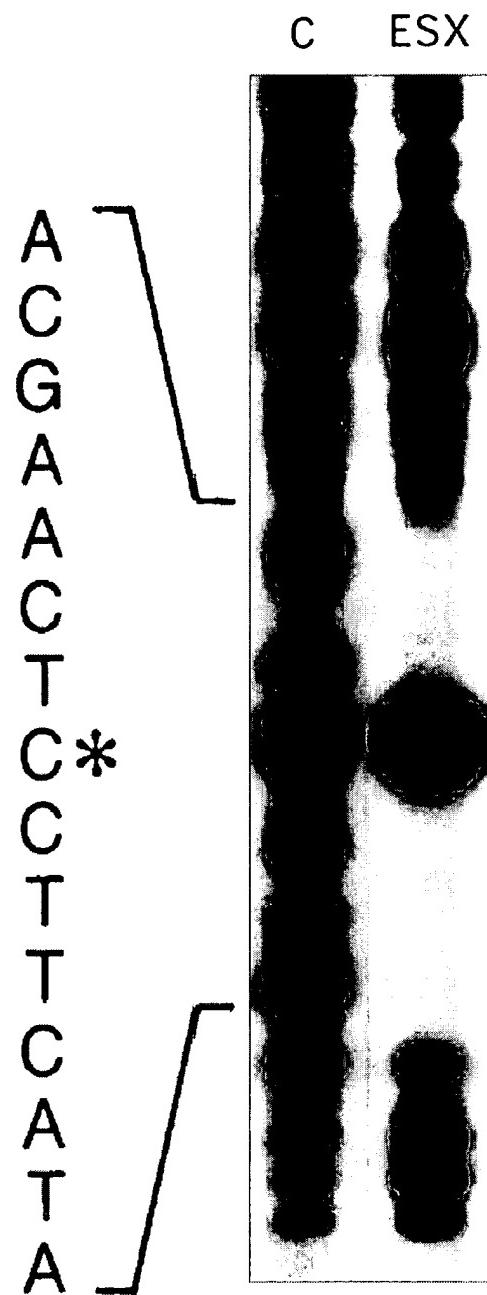


Fig. 8B

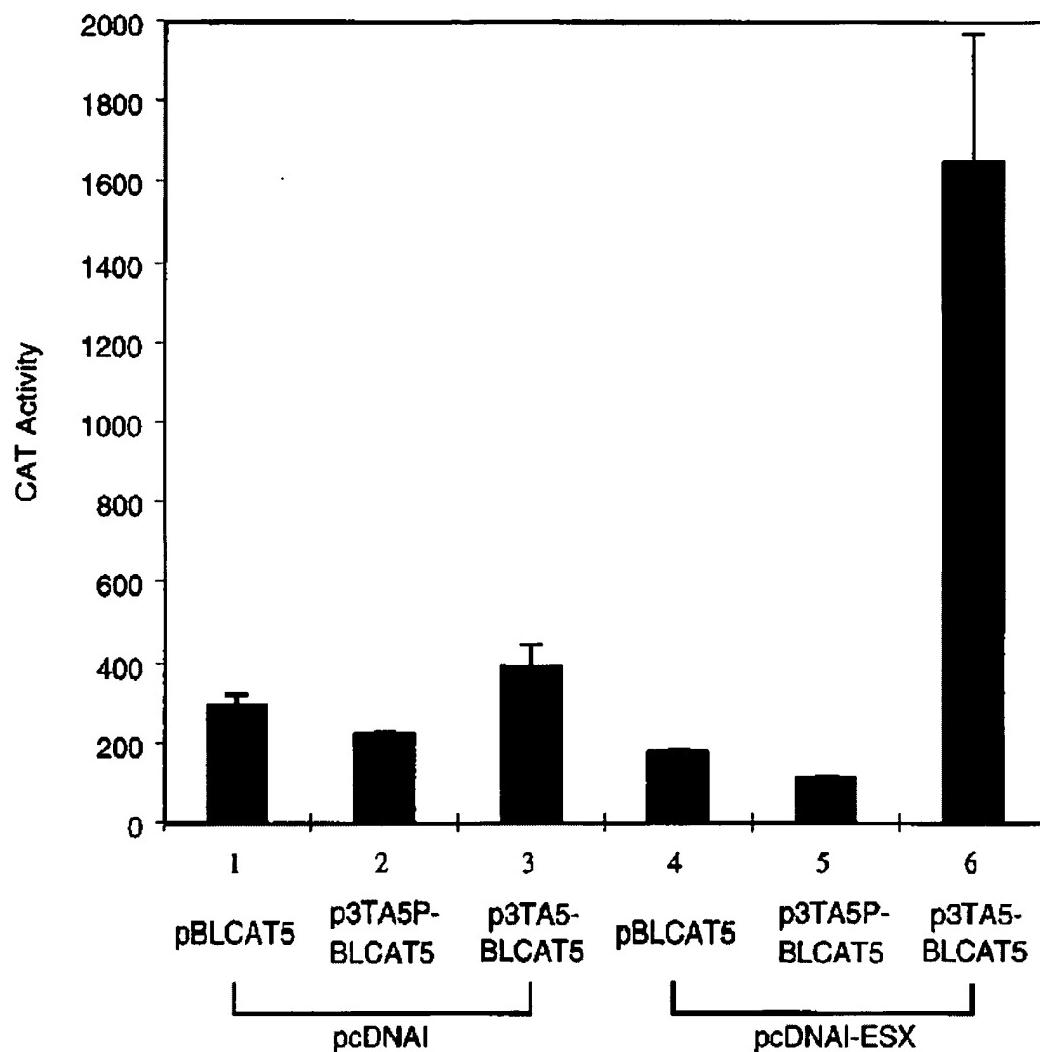


Fig. 8C

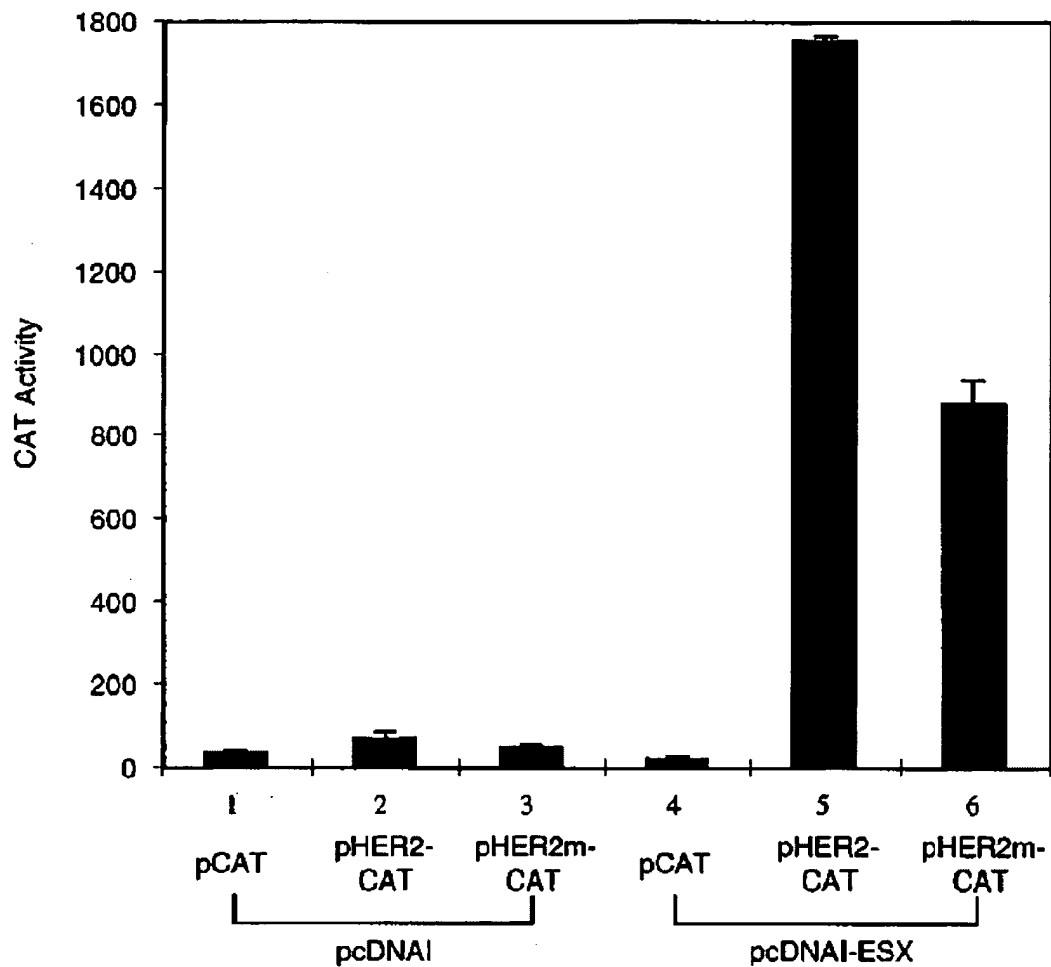


Fig. 8D

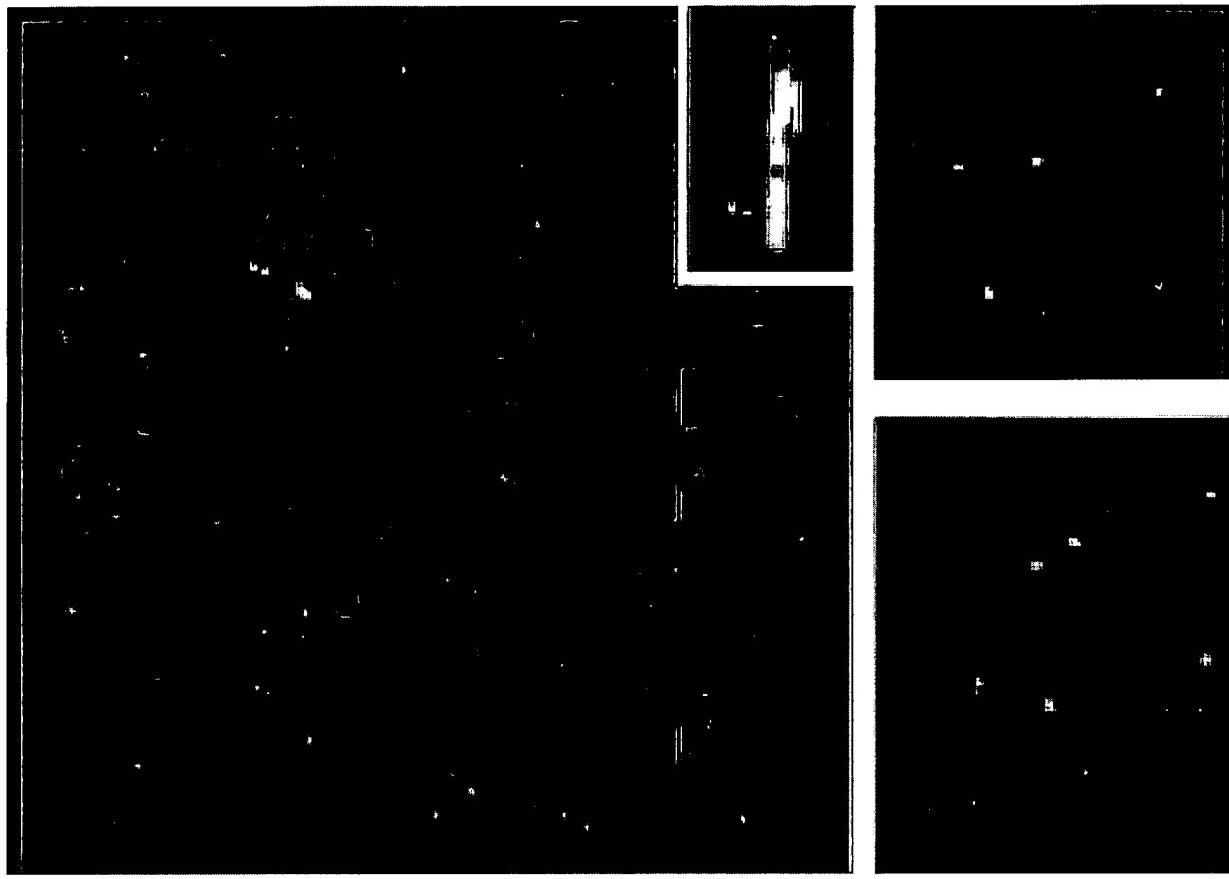


Fig. 8E

RECEIVED
U.S. PATENT & TRADEMARK OFFICE

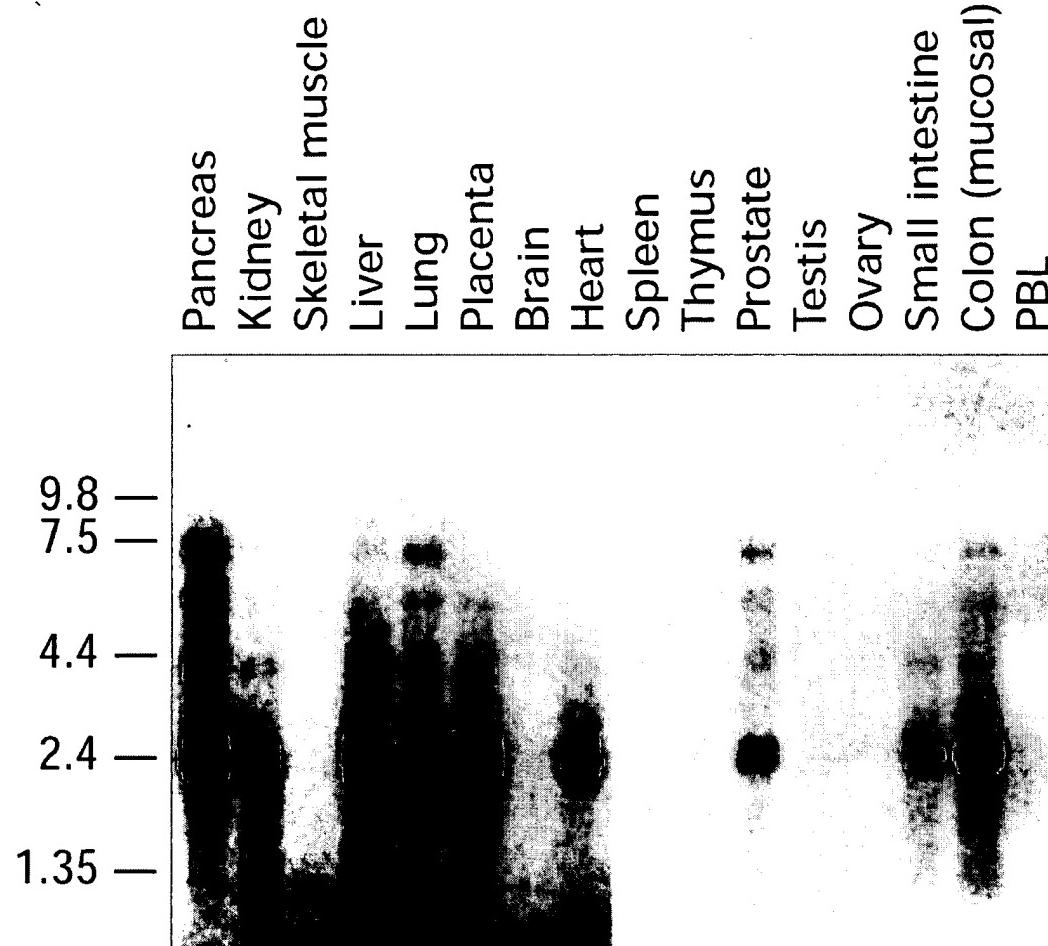


Fig. 9A

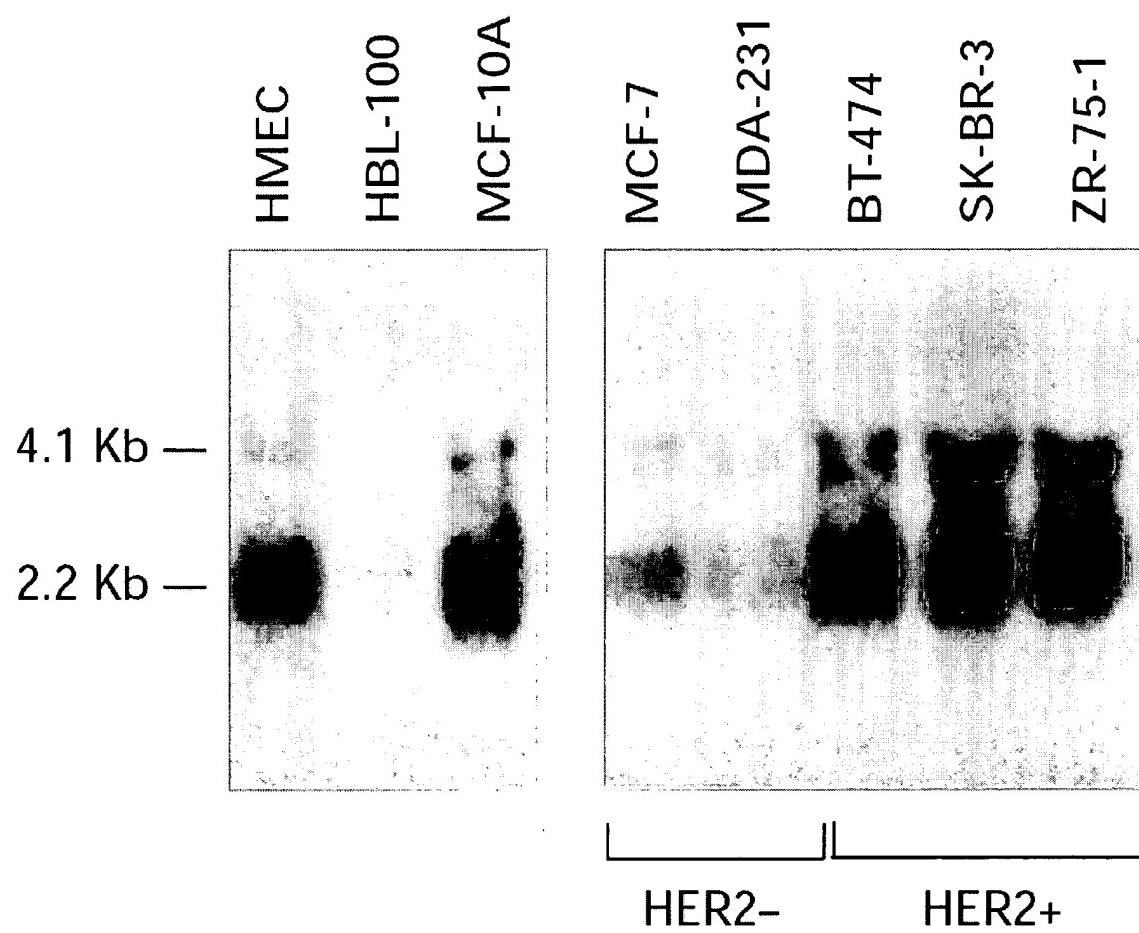


Fig. 9B

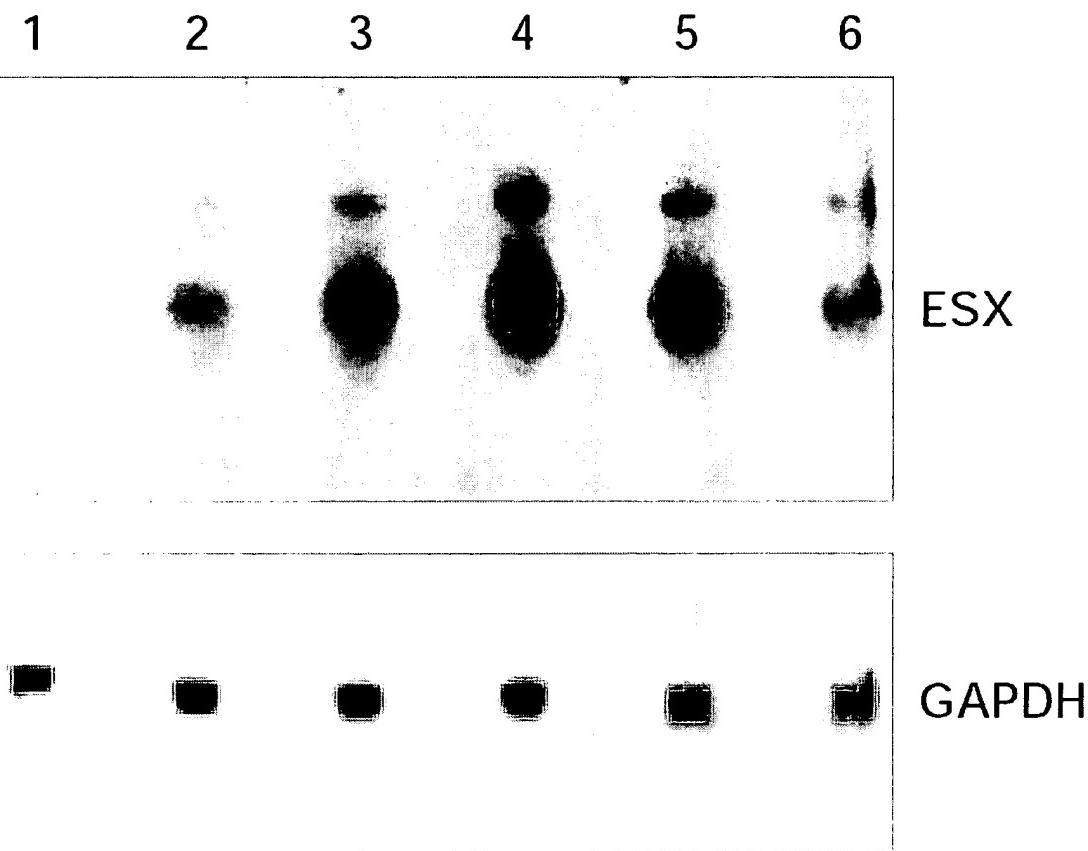


Fig. 9C

PRINTED IN U.S.A.
U.S. GOVERNMENT PRINTING OFFICE
1938 10 10 1938

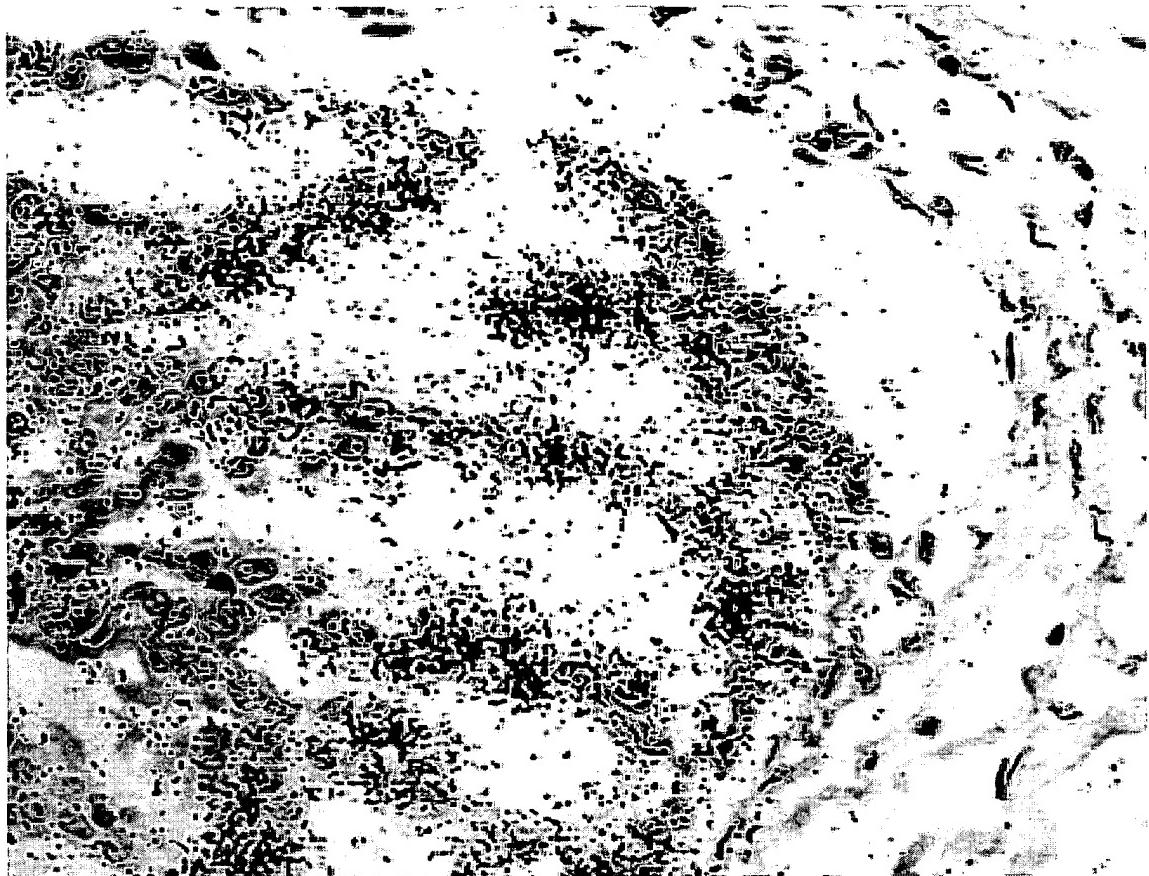


Fig. 10A

CONFIDENTIAL
DO NOT DISTRIBUTE

U.S. PATENT & TRADEMARK OFFICE
NOV 19 2002

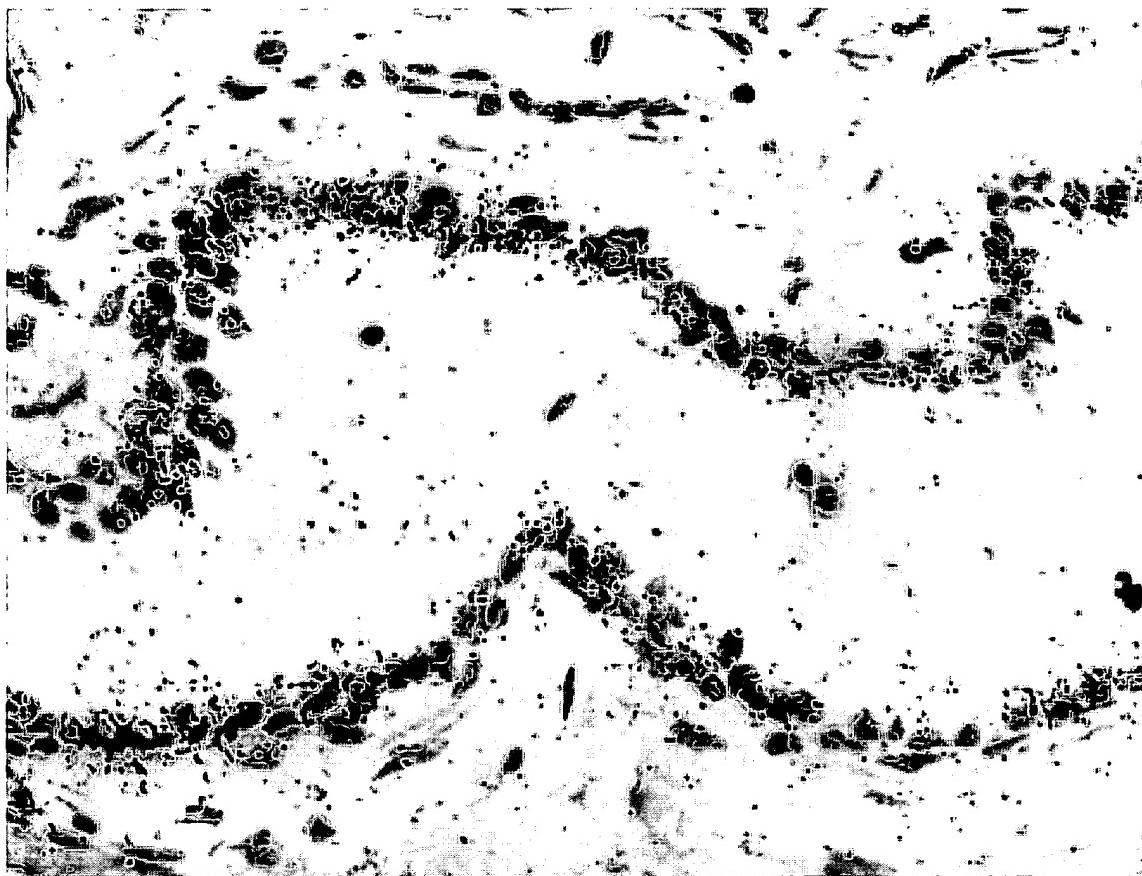


Fig. 10B

SEARCHED
SERIALIZED
INDEXED
FILED
APR 22 2003
U.S. PATENT & TRADEMARK OFFICE